A Beta-Beta Achievability Bound with Applications

A Beta-Beta Achievability Bound with Applications

Wei Yang[1], Austin Collins[2], Giuseppe Durisi[3],
Yury Polyanskiy[2], and H. Vincent Poor[1]
[1]Princeton University, Princeton, NJ, 08544 USA
[2]Massachusetts Institute of Technology, Cambridge, MA, 02139 USA
[3]Chalmers University of Technology, 41296 Gothenburg, Sweden

***Abstract*—A channel coding achievability bound expressed in terms of the ratio between two Neyman-Pearson β functions is proposed. This bound is the dual of a converse bound established earlier by Polyanskiy and Verdú (2014). The new bound turns out to simplify considerably the analysis in situations where the channel output distribution is not a product distribution, for example due to a cost constraint or a structural constraint (such as orthogonality or constant composition) on the channel inputs. Connections to existing bounds in the literature are discussed. The bound is then used to derive 1) the channel dispersion of additive non-Gaussian noise channels with random Gaussian codebooks, 2) the channel dispersion of an exponential-noise channel, 3) a second-order expansion for the minimum energy per bit of an AWGN channel, and 4) a lower bound on the maximum coding rate of a multiple-input multiple-output Rayleigh-fading channel with perfect channel state information at the receiver, which is the tightest known achievability result.**

I. Introduction

We consider an abstract channel that consists of an input set $\mathcal{A}$, an output set $\mathcal{B}$, and a random transformation $P_{Y|X}: \mathcal{A} \to \mathcal{B}$. An (M, ϵ) code for the channel $(\mathcal{A}, P_{Y|X}, \mathcal{B})$ comprises a message set $\mathcal{M} \triangleq \{1, \ldots, M\}$, an encoder $f: \mathcal{M} \to \mathcal{A}$, and a decoder $g: \mathcal{B} \to \mathcal{M} \cup \{e\}$ (e denotes an error event) that satisfies the *average* error probability constraint

$$\frac{1}{M}\sum_{j=1}^{M}\Big(1 - P_{Y\,|\,X}\big(g^{-1}(j)\,|\,f(j)\big)\Big) \le \epsilon. \tag{1}$$

Here, $g^{-1}(j) \triangleq \{y \in \mathcal{Y} : g(y) = j\}$. For a fixed arbitrary $\epsilon \in (0,1)$, we are interested in finding a lower bound (i.e., an achievability bound) on the largest number M^* of codewords for which an (M, ϵ) code exists.

For stationary memoryless channels, Shannon's channel coding theorem establishes that the rate of the best code converges to the channel capacity

$$C = \max_{P_X} I(X;Y) \tag{2}$$

as the blocklength grows to infinity. Here, $I(X;Y)$ denotes the mutual information between the channel input and output. The mutual information can be expressed through an arbitrary output distribution Q_Y as follows [1, Eq. (8.7)]:

$$I(X;Y) = D(P_{Y|X}\|Q_Y|P_X) - D(P_Y\|Q_Y). \tag{3}$$

This work was supported in part by the US National Science Foundation (NSF) under Grants CCF-1420575 and ECCS-1343210, by the Swedish Research Council, under grant 3222452, by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, and by the NSF CAREER award CCF-12-53205.

This identity—also known as the *golden formula*—has found many applications in information theory. For example, it allows us to prove upper bounds on channel capacity (by dropping the term $-D(P_Y\|Q_Y)$; see [2]). It is also used in the derivation of the capacity per unit cost [3], in the Blahut-Arimoto algorithm [4], [5], in Gallager's formula for the minimax redundancy in universal source coding [6], and in characterizing properties of good channel codes [7], [8].

As a first step, Polyanskiy and Verdú recently proved that every (M, ϵ) code satisfies the following converse bound [8, Th. 15]

$$M \le \inf_{0\le\delta<1-\epsilon}\inf_{Q_Y}\frac{\beta_{1-\delta}(P_Y, Q_Y)}{\beta_{1-\epsilon-\delta}(P_{XY}, P_XQ_Y)}. \tag{4}$$

Here, P_X and P_Y denote the empirical input and output distributions induced by the code (for the case of uniformly distributed messages). The function $\beta_\alpha(P,Q)$ in (4) for two probability measures P and Q on $\mathcal{X}$ measures the difficulty of distinguishing P from Q in terms of hypothesis testing, and is defined as[1]

$$\beta_\alpha(P,Q) \triangleq \min \int P_{Z\,|\,X}(1\,|\,x)Q(dx) \tag{5}$$

where the minimum is over all conditional probability distributions (i.e., tests) $P_{Z\,|\,X}: \mathcal{X} \to \{0,1\}$ satisfying

$$\int P_{Z\,|\,X}(1\,|\,x)P(dx) \ge \alpha. \tag{6}$$

The analogy between (3) and (4) follows from Stein's lemma:

$$-\log\beta_\alpha(P^n, Q^n) = nD(P\|Q) + o(n), \quad \forall \alpha \in (0,1). \tag{7}$$

Contributions: In this paper, we continue the program of establishing a finite-blocklength analog of the golden formula by proving the following achievability counterpart of (4).

Theorem 1 ($\beta\beta$ achievability bound): For every $0 < \epsilon < 1$ and every input distribution P_X, there exists an (M, ϵ) code for the channel $(\mathcal{A}, P_{Y|X}, \mathcal{B})$ satisfying

$$\frac{M}{2} \ge \sup_{0<\tau<\epsilon}\sup_{Q_Y}\frac{\beta_\tau(P_Y, Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY}, P_XQ_Y)} \tag{8}$$

where $P_Y \triangleq P_{Y|X} \circ P_X$.

The proof of this bound relies on Shannon's random coding technique and on a suboptimal decoder that is based on the

[1]By the Neyman-Pearson lemma [9], there exists an optimal $P_{Z|X}$ that attains the minimum in (5). This test, which we shall refer to as the Neyman-Pearson test, involves thresholding the Radon-Nikodym derivative of P with respect to Q.

Neyman-Pearson test between P_{XY} and $P_X Q_Y$. Hypothesis testing is used twice in the proof: to relate the decoding error probability to $\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)$, and to perform a change of measure from P_Y to Q_Y.

The bound (8) is useful in situations where P_Y is not a product distribution (although the underlying channel law $P_{Y|X}$ is stationary and memoryless), for example due to cost constraints, or structural constraints on the channel input, such as orthogonality or constant composition. In such cases, traditional achievability bounds such as Feinstein's bound [10] and the dependence-testing (DT) bound [11, Th. 18], which are explicit in $dP_{Y|X}/dP_Y$, become difficult to evaluate. In contrast, the $\beta\beta$ bound (8) requires the evaluation of $dP_{Y|X}/dQ_Y$, which factorizes for product Q_Y. This allows for an analytical computation of (8). Furthermore, the term $\beta_\tau(P_Y, Q_Y)$, which captures the cost of the change of measure from P_Y to Q_Y, can be evaluated or bounded even when a closed-form expression for P_Y is not available. To illustrate these points, we present the following applications of Theorem 1:

- We obtain the channel dispersion [11, Def. 1] of additive non-Gaussian noise channels, for the case in which the encoder uses a power-constrained random Gaussian codebook. We show that the power constraint introduces an additional term in the expression of the achievable dispersion, which depends on the minimum mean square error (MMSE) of estimating the channel input given the channel output.
- We characterize the channel dispersion of the additive exponential noise channel introduced in [12]. The channel dispersion of a discrete couterpart of the exponential-noise channel is studied in [13].
- We prove a second-order expansion for the minimum energy per bit of an additive white Gaussian noise (AWGN) channel at finite blocklength, hence establishing a nonasymptotic counterpart of the *wideband slope* result of Verdú [14]. Even though this result can be obtained via other techniques (such as the $\kappa\beta$ bound [11, Th. 25]), the proof based on (8) is conceptually simpler and generalizes to other channel models. Furthermore, the converse part of this result is proved using the $\beta\beta$ converse bound (4).
- We evaluate (8) for a multiple-input multiple-output (MIMO) Rayleigh-fading channel with perfect channel state information at the receiver (CSIR). In this case, (8) yields the tightest known achievability result.

Notation: For an input distribution P_X and a channel $P_{Y|X}$, we let $P_{Y|X} \circ P_X$ denote the distribution of Y induced by P_X through $P_{Y|X}$. The distribution of a circularly symmetric complex Gaussian random vector with covariance matrix A is denoted by $\mathcal{CN}(0, \mathsf{A})$. With $\chi^2_k(\lambda)$ we denote the noncentral chi-sqared distribution with k degrees of freedom and noncentrality parameter λ. Finally, $\mathrm{Exp}(\mu)$ stands for the exponential distribution with mean μ.

II. Proof of Theorem 1

Fix $\epsilon \in (0,1)$, $\tau \in (0, \epsilon)$, and let P_X and Q_Y be two arbitrary probability measures on $\mathcal{A}$ and $\mathcal{B}$, respectively. Furthermore, let

$$M = \left\lceil \frac{2\beta_\tau(P_Y, Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)} \right\rceil. \tag{9}$$

Finally, let $P_{Z\,|\,X,Y} : \mathcal{A} \times \mathcal{B} \to \{0,1\}$ be the Neyman-Pearson test that satisfies

$$P_{XY}[Z(X,Y) = 1] \geq 1 - \epsilon + \tau \tag{10}$$

$$P_X Q_Y[Z(X,Y) = 1] = \beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y). \tag{11}$$

For a given codebook $\{c_1, \ldots, c_M\}$ and a received signal y, the decoder computes the values of $Z(c_j, y)$ and returns the smallest index j for which $Z(c_j, y) = 1$. If no such index is found, the decoder declares an error. The average probability of error of the given codebook $\{c_1, \ldots, c_M\}$, under the assumption of uniformly distributed messages, is given by

$$P_{\mathrm{e}}(c_1, \ldots, c_M) = \mathbb{P}\left[\big\{Z(c_W, Y) = 0\big\} \bigcup_{m<W} \big\{Z(c_m, Y) = 1\big\}\right] \tag{12}$$

where W is equiprobable on $\{1, \ldots, M\}$ and $Y \sim P_{Y|W}$.

Following Shannon's random coding idea, we next average (12) over all codebooks $\{C_1, \ldots, C_M\}$ whose codewords are generated as pairwise independent random variables with distribution P_X. This yields

$$\begin{aligned}
&\mathbb{E}[P_{\mathrm{e}}(C_1, \ldots, C_M)] \\
&\quad \leq \mathbb{P}\big[Z(X,Y) = 0\big] + \mathbb{P}\left[\max_{m<W} Z(C_m, Y) = 1\right] && (13)\\
&\quad \leq \epsilon - \tau + \mathbb{P}\left[\max_{m<W} Z(C_m, Y) = 1\right]. && (14)
\end{aligned}$$

Here, (13) follows from the union bound and (14).

To conclude the proof of (8), it suffices to show that

$$\mathbb{P}\left[\max_{m<W} Z(C_m, Y) = 1\right] \leq \tau. \tag{15}$$

Consider the randomized test $P_{\widetilde{Z}\,|\,Y} : \mathcal{Y} \to \{0,1\}$:

$$\widetilde{Z}(y) \triangleq \max_{m<W} Z(C_m, y). \tag{16}$$

It follows that

$$\begin{aligned}
\beta_{P_Y[\widetilde{Z}=1]}(P_Y, Q_Y) &\leq Q_Y[\widetilde{Z}(Y) = 1] && (17)\\
&\leq \frac{1}{M}\sum_{j=1}^{M}(j-1) P_X Q_Y[Z(X,Y) = 1] && (18)\\
&= \frac{M-1}{2} P_X Q_Y[Z(X,Y) = 1] && (19)\\
&= \frac{M-1}{2}\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y) && (20)\\
&\leq \beta_\tau(P_Y, Q_Y). && (21)
\end{aligned}$$

Here, (17) follows from (5); (18) follows from (16) and from the union bound; (20) follows from (11); and (21) follows from (9). Since $\alpha \mapsto \beta_\alpha(P_Y, Q_Y)$ is nondecreasing, we conclude that

$$P_Y[\widetilde{Z} = 1] \leq \tau \tag{22}$$

which is equivalent to (15).

III. Connection to Existing Bounds

We next illustrate the connection between Theorem 1 and other achievability bounds.

1) The $\kappa\beta$ bound [11, Th. 25]: The $\kappa\beta$ bound is based on Feinstein's maximal coding approach and on a suboptimal decoder similar to the one used in Theorem 1. By further lower-bounding the κ term in the $\kappa\beta$ bound using [15, Lemma 4], we can relax it to the following bound:

$$M \geq \sup_{\tau\in(0,\epsilon)} \sup_{Q_Y} \frac{\beta_\tau(P_{Y|X}\circ P_X, Q_Y)}{\sup_{x\in\mathcal{F}} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)} \tag{23}$$

which holds under a *maximum* error probability constraint. Here, $\mathcal{F}\subset\mathcal{A}$ denotes the permissible set of codewords, and P_X is an arbitrary distribution on $\mathcal{F}$. The similarity between (23) and (8) suggests that we can interpret the $\beta\beta$ bound as the average-error-probability counterpart of the $\kappa\beta$ bound. For the case in which $\beta_\alpha(P_{Y|X=x}, Q_Y)$ does not depend on $x\in\mathcal{F}$, by relaxing $M/2$ to M in (8) and by using [11, Lemma 29] we obtain a weaker version of (23) that holds under the average error probability constraint. However, for the case in which $\beta_\alpha(P_{Y|X=x}, Q_Y)$ does depend on $x\in\mathcal{F}$, (8) can be both easier to analyze and numerically tighter than (23) (see Section IV-D for an example).

2) The dependence-testing (DT) bound [11, Th. 18]: Setting $Q_Y = P_Y$ in (8), using that $\beta_\tau(P_Y, P_Y) = \tau$, and rearranging terms, we obtain

$$\epsilon \leq \inf_{\alpha\in(0,1)} \left\{1-\alpha+\frac{M}{2}\beta_\alpha(P_{XY}, P_XP_Y)\right\}. \tag{24}$$

Setting $\alpha = P_{XY}[\log\left(dP_{XY}/d(P_XP_Y)\right) \geq \log(M/2)]$ and using the Neyman-Pearson lemma, we conclude that (24) is equivalent to a slightly weakened version of the DT bound with $(M-1)/2$ replaced by $M/2$. Since this weakened version of the DT bound implies Shannon's bound [16] and the bound in [17, Th. 2], our bound implies these two bounds as well.

IV. Applications

We shall take $\mathcal{A}$ and $\mathcal{B}$ to be n-fold Cartesian products of alphabets $\mathcal{X}$ and $\mathcal{Y}$. A channel is a sequence of conditional probabilities $P_{Y^n|X^n}: \mathcal{X}^n \to \mathcal{Y}^n$. We shall refer to an (M,ϵ) code for the channel $\{\mathcal{X}^n, P_{Y^n|X^n}, \mathcal{Y}^n\}$ as an (n,M,ϵ) code. Furthermore, the maximum coding rate $R^*(n,\epsilon)$ is defined as[2]

$$R^*(n,\epsilon) \triangleq \sup\left\{\frac{\log M}{n} : \exists(n,M,\epsilon)\text{ code}\right\}. \tag{25}$$

Due to space limitations, we have omitted the proofs of all theorems in the section. They can be found in [18].

A. Additive non-Gaussian noise channels

We consider the additive-noise channel

$$Y_i = X_i + Z_i, \quad i=1,\dots,n \tag{26}$$

where $\{Z_i\}$ are independent and identically distributed (i.i.d.) P_Z-distributed (not necessarily Gaussian) and $X_i, Y_i, Z_i \in \mathbb{R}$. Each codeword x^n must satisfy the constraint

$$\|x^n\|^2 = \sum_{i=1}^n x_i^2 \leq nP. \tag{27}$$

Let $Q_{X^n} = \mathcal{N}(\mathbf{0}, P\mathsf{I}_n)$, and let P_{X^n} denote the conditional distribution of $X^n \sim Q_{X^n}$ conditioned on

$$X^n \in \mathcal{A}_n \triangleq \left\{x^n\in\mathbb{R}^n : nP - \log n \leq \|x^n\|^2 \leq nP\right\}. \tag{28}$$

In other words, P_{X^n} is a truncated Gaussian distribution that is supported on the spherical shell $\mathcal{A}_n$. We shall consider an ensemble of codes in which the codewords are generated independently from the distribution P_{X^n}. This ensemble of codes is used by Gallager to derive the random coding error exponent for channels with cost constraint [19, p. 326]. Let $R^*_{\mathrm{G}}(n,\epsilon)$ be the maximum achievable rate using this ensemble of codes over the channel (26) with average error probability $\mathbb{P}[W\neq\hat{W}]\leq\epsilon$, where the probability is over the equiprobable message W, the random codebook ensemble and the channel noise $\{Z_i\}$. In the following theorem we present an ensemble-tight second-order asymptotic expansion for $R^*_{\mathrm{G}}(n,\epsilon)$.

Theorem 2: Let $Q_X = \mathcal{N}(0,P)$, let $Q_Y = P_{Y|X}\circ Q_X$, and denote the information density of $Q_XP_{Y|X}$ by

$$i(x;y) \triangleq \frac{dP_{Y|X}}{dQ_Y}(x;y). \tag{29}$$

Furthermore, let

$$I(P) \triangleq \mathbb{E}_{Q_XP_{Y|X}}[i(X;Y)]. \tag{30}$$

Assume that the noise Z satisfies the following conditions:

1) P_Z is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$;
2) $\mathbb{E}_{Q_XP_Z}\left[|i(X;X+Z)-I(P)|^3\right] < \infty$; and
3) $\mathbb{E}\left[|Z|^6\right] < \infty$.

Then, for every $0<\epsilon<1$, we have

$$R^*_{\mathrm{G}}(n,\epsilon) = I(P) - \sqrt{\frac{V(P)}{n}}Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right). \tag{31}$$

Here, $Q^{-1}(\cdot)$ denotes the inverse of the Gaussian Q-function,

$$V(P) \triangleq \mathrm{Var}[i(X;Y) + c(X^2-P)] \tag{32}$$

where

$$c \triangleq \frac{\log e}{2P^2}\Big(\mathsf{mmse}(X|Y) - P\Big) \tag{33}$$

and

$$\mathsf{mmse}(X|Y) \triangleq \mathbb{E}\left[(X-\mathbb{E}[X|Y])^2\right]. \tag{34}$$

In (32) and (34), the pair (X,Y) is distributed according to $Q_XP_{Y|X}$.

Remark 1: For $P_Z = \mathcal{N}(0,1)$, (32) recovers the dispersion $V(P) = \frac{P(2+P)}{2(1+P)^2}\log^2 e$ of the AWGN channel [11, Th. 54].

Remark 2: By removing the codewords $\boldsymbol{x}$ for which $\left|\sum_{i=1}^n D(P_{Y|X=x_i}\|Q_Y) - nI(P)\right| \gg 1$ from $\mathcal{A}_n$, it is possible to achieve a dispersion that is equal to $\mathrm{Var}[i(X;Y)|X]$, provided that the noise distribution P_Z satisfies further regularity conditions; see [20, Th. 5 and Section VI].

[2]Unless otherwise indicated, the log and exp functions in this paper are taken with respect to an arbitrary fixed basis.

B. The exponential-noise channel

We next consider the exponential-noise case, i.e., $P_Z = \mathrm{Exp}(1)$. As in [12], we assume that each codeword $x^n \in \mathbb{R}^n$ must satisfy

$$x_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} x_i \leq n\sigma. \tag{35}$$

The practical relevance of such a channel is discussed in [12], [21]. The capacity of the exponential-noise channel with constraint (35) is given by [12, Th. 3]

$$C(\sigma) = \log(1+\sigma) \tag{36}$$

and is achieved by the input distribution P_X^*, according to which X takes the value zero with probability $1/(1+\sigma)$ and follows an $\mathrm{Exp}(1+\sigma)$ distribution conditioned on it being positive. Furthermore, the capacity-achieving output distribution is $\mathrm{Exp}(1+\sigma)$.

Theorem 3: For the additive exponential-noise channel subject to the constraint (35) and for $0 < \epsilon < 1$,

$$R^*(n,\epsilon) = \log(1+\sigma) - \sqrt{\frac{V(\sigma)}{n}}Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \tag{37}$$

where

$$V(\sigma) = \frac{\sigma^2}{(1+\sigma)^2}\log^2 e. \tag{38}$$

C. Minimum energy per bit over AWGN channels

For a complex-valued AWGN channel, we set $\mathcal{A} = \mathbb{C}^n$, $\mathcal{B} = \mathbb{C}^n$, and $P_{Y^n|X^n=x^n} = \mathcal{CN}(x^n, \mathsf{I}_n)$. We assume that every codeword x^n satisfies the equal power constraint

$$\|x^n\|^2 = nP. \tag{39}$$

Let $R_{\mathrm{e}}^*(n,\epsilon,P)$ denote the maximum coding rate $R^*(n,\epsilon)$ under the constraint (39). Theorem 4 below provides expressions for the β functions in (4) and (8) for the AWGN case.

Theorem 4: Consider the complex-valued AWGN channel $P_{Y^n|X^n}$. Let $S_n \sim \chi^2_{2n}(2nP)$ and $L_n \sim \chi^2_{2n}(0)$. Let $Q_{Y^n} = \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$. Furthermore, let $\mathcal{S}_n \triangleq \{x^n \in \mathbb{C}^n : \|x^n\|^2 = nP\}$. Then, for every distribution P_{X^n} supported on $\mathcal{S}_n$

$$\beta_\alpha(P_{X^nY^n}, P_{X^n}Q_{Y^n}) = Q\left(\sqrt{2nP} + Q^{-1}(\alpha)\right) \tag{40}$$

and

$$\beta_a(P_{Y^n|X^n} \circ P_{X^n}, Q_{Y^n}) \leq \mathbb{P}[L_n \geq \gamma] \tag{41}$$

where γ satisfies

$$\mathbb{P}[S_n \geq \gamma] = a. \tag{42}$$

Furthermore, (41) holds with equality if P_{X^n} is the uniform distribution over $\mathcal{S}_n$.

By evaluating (40) and (41) in the asymptotic regime $P \to 0$ and $nP^2 \to \infty$ as $n \to \infty$,[3] and by substituting them in Theorem 1 and in (4), we obtain the following result.

[3]As we shall see, this regime is of interest for the characterization of the minimum energy per bit.

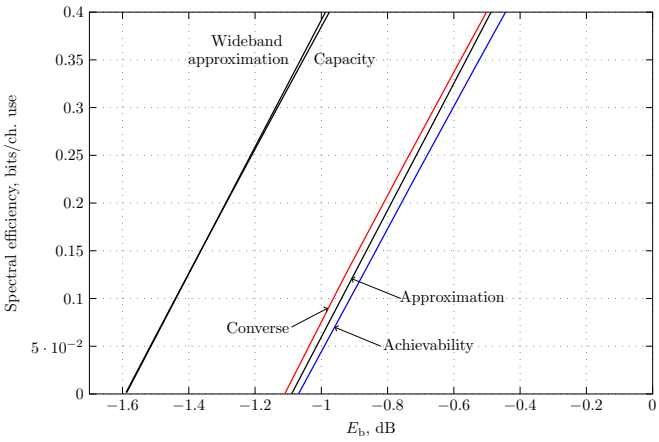

Figure 1. Minimum energy per bit versus spectral efficiency of the AWGN channel; here, $k = 2000$ bits, and $\epsilon = 10^{-3}$.

Theorem 5: For an AWGN channel with SNR P_n satisfying $P_n \to 0$ and $nP_n^2 \to \infty$ as $n \to \infty$, the maximum coding rate $R_{\mathrm{e}}^*(n,\epsilon,P_n)$ behaves as

$$\frac{R_{\mathrm{e}}^*(n,\epsilon,P_n)}{\log e} = P_n - \sqrt{\frac{2P_n}{n}}Q^{-1}(\epsilon) - \frac{1}{2}P_n^2 + o\left(\sqrt{\frac{P_n}{n}}\right) + o\left(P_n^2\right), \quad n \to \infty. \tag{43}$$

We now relate (43) to the minimum energy per bit $E_{\mathrm{b}}^*(k,\epsilon,R)$ to transmit k information bits at rate R and error probability ϵ. Specifically, Theorem 5 implies that

$$10\log_{10} E_{\mathrm{b}}^*(k,\epsilon,R)$$
$$= 10\log_{10}\frac{P_n}{R_{\mathrm{e}}^*(n,\epsilon,P_n)} \tag{44}$$
$$= 10\log_{10}\left(\log_e 2 + \sqrt{\frac{2\log_e 2}{k}}Q^{-1}(\epsilon) + \frac{\log_e^2 2}{2}R\right) + o(R) + o(1/\sqrt{k}) \tag{45}$$
$$= 10\log_{10}E_{\mathrm{b}}^*(k,\epsilon,0) + \frac{10\log_{10}2}{2}R + o(R) + o\left(\frac{1}{\sqrt{k}}\right). \tag{46}$$

The last step follows from [22, Th. 3]. Note that (46) is the finite-blocklength counterpart of Verdú's *wideband-slope* result [14, Eq. (172)].

In Fig. 1, we present a comparison between the approximation (46) (with the $o(\cdot)$ terms omitted), the converse bound [11, Th. 28], and the achievability bound (8). In both cases Q_Y is chosen to be the capacity-achieving output distribution. For the parameters considered in Fig. 1, the approximation (46) is accurate.

D. MIMO block-fading channel with perfect CSIR

Consider an $m_{\mathrm{t}} \times m_{\mathrm{r}}$ Rayleigh MIMO block-fading channel that stays constant for n_{c} channel uses. The input-output relation within the kth coherence interval is given by

$$\mathbb{Y}_k = \mathbb{X}_k\mathbb{H}_k + \mathbb{W}_k. \tag{47}$$

Here, $\mathbb{X}_k \in \mathbb{C}^{n_{\mathrm{c}}\times m_{\mathrm{t}}}$ and $\mathbb{Y}_k \in \mathbb{C}^{n_{\mathrm{c}}\times m_{\mathrm{r}}}$ are the transmitted and received matrices, respectively; the entries of the fading

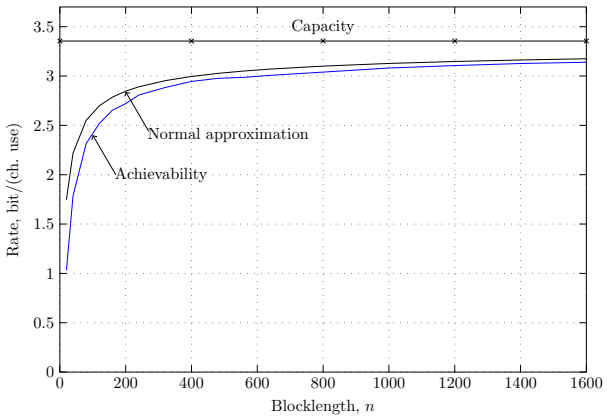

Figure 2. Bounds on rate for a 4×4 MIMO Rayleigh block-fading channel; here SNR=0 dB, $\epsilon = 0.001$, and $n_{\mathrm{c}} = 4$.

matrix $\mathbb{H}_k \in \mathbb{C}^{m_{\mathrm{t}} \times m_{\mathrm{r}}}$ and the noise $\mathbb{W}_k \in \mathbb{C}^{n_{\mathrm{c}} \times m_{\mathrm{r}}}$ are i.i.d. $\mathcal{CN}(0,1)$. We assume that $\{\mathbb{H}_k\}$ and $\{\mathbb{W}_k\}$ take on independent realizations over successive coherence intervals. The channel matrices $\{\mathbb{H}_k\}$ are assumed to be known to the receiver but not to the transmitter. We shall also assume that each codeword spans $l \in \mathbb{N}$ coherence intervals, i.e., the blocklength of the code is $n = ln_{\mathrm{c}}$. Finally, each codeword X^l is constrained to satisfy

$$\left\|\mathsf{X}^l\right\|_{\mathsf{F}} \leq \sqrt{nP}. \tag{48}$$

To obtain an achievability bound on $R^*(n,\epsilon)$, we apply Theorem 1 with $P_{\mathbb{X}^l}$ chosen as the uniform distribution on $\mathcal{S}'_n \triangleq \{\mathsf{X}^l : \left\|\mathsf{X}^l\right\|_{\mathsf{F}}^2 = nP\}$ and $Q_{\mathbb{Y}^l\mathbb{H}^l}$ chosen as the capacity-achieving output distribution. With these choices, we have

$$R^*(n,\epsilon) \geq \frac{1}{n} \log \frac{\beta_\tau(P_{\mathbb{H}^l\mathbb{Y}^l}, Q_{\mathbb{H}^l\mathbb{Y}^l})}{\beta_{1-\epsilon+\tau}(P_{\mathbb{X}^l\mathbb{H}^l\mathbb{Y}^l}, P_{\mathbb{X}^l}Q_{\mathbb{H}^l\mathbb{Y}^l})}. \tag{49}$$

The denominator $\beta_{1-\epsilon+\tau}(P_{\mathbb{X}^l\mathbb{H}^l\mathbb{Y}^l}, P_{\mathbb{X}^l}Q_{\mathbb{H}^l\mathbb{Y}^l})$ in (49) can be computed via standard Monte Carlo techniques. However, computing $\beta_\tau(P_{\mathbb{H}^l\mathbb{Y}^l}, Q_{\mathbb{H}^l\mathbb{Y}^l})$ in the numerator is more involved, since there is no closed-form expression for $P_{\mathbb{H}^l\mathbb{Y}^l}$. To circumvent this, we further lower-bound $\beta_\tau(P_{\mathbb{H}^l\mathbb{Y}^l}, Q_{\mathbb{H}^l\mathbb{Y}^l})$ using the data-processing inequality [23] for β_α as follows. Let $\widetilde{\mathbb{X}}^l$ be a sequence of $n_{\mathrm{c}} \times m_{\mathrm{t}}$ complex matrices with i.i.d. $\mathcal{CN}(0, P/m_{\mathrm{t}})$ entries. Then, $P_{\mathbb{X}^l}$ can be obtained via $\widetilde{\mathbb{X}}^l$ through $\mathbb{X}^l = \sqrt{nP}\widetilde{\mathbb{X}}^l / \left\|\widetilde{\mathbb{X}}^l\right\|_{\mathsf{F}}$. Furthermore, $Q_{\mathbb{H}^l\mathbb{Y}^l} = P_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l} \circ P_{\widetilde{\mathbb{X}}^l}$. Let $P^{(s)}_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l} \triangleq P_{\mathbb{H}^l} P^{(s)}_{\mathbb{Y}^l \,|\, \mathbb{H}^l, \mathbb{X}^l}$, where $P^{(s)}_{\mathbb{Y}^l \,|\, \mathbb{H}^l, \mathbb{X}^l}$ denotes the channel law defined by

$$\mathbb{Y}_k = \mathbb{X}_k \mathbb{H}_k \frac{\sqrt{nP}}{\left\|\mathbb{X}^l\right\|_{\mathsf{F}}} + \mathbb{W}_k, \qquad k = 1, \ldots, l. \tag{50}$$

We have that $P_{\mathbb{Y}^l\mathbb{H}^l} = P_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l} \circ P_{\mathbb{X}^l} = P^{(s)}_{\mathbb{Y}^l\mathbb{H}^l \,|\, f\mathbb{X}^l} \circ P_{\widetilde{\mathbb{X}}^l}$. Now, by the data-processing inequality,

$$\beta_\tau(P_{\mathbb{H}^l\mathbb{Y}^l}, Q_{\mathbb{H}^l\mathbb{Y}^l}) \geq \beta_\tau(P_{\widetilde{\mathbb{X}}^l} P^{(s)}_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l}, P_{\widetilde{\mathbb{X}}^l} P_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l}). \tag{51}$$

Since the Radon-Nikodym derivative $\frac{d(P_{\widetilde{\mathbb{X}}^l} P^{(s)}_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l})}{d(P_{\widetilde{\mathbb{X}}^l} P_{\mathbb{Y}^l\mathbb{H}^l \,|\, \mathbb{X}^l})}$ can be computed in closed form, the right-hand side of (51) can be computed via Monte Carlo techniques. The resulting bound is compared with the normal approximation of $R^*(n,\epsilon)$ in Fig. 2. In contrast, the $\kappa\beta$ bound [11, Th. 25] with $\mathcal{F} = \mathcal{S}'_n$ is much more difficult to compute due to the maximization over codewords $\mathsf{X}^l \in \mathcal{S}'_n$. Furthermore, for blocklength values of practical interest, we expect that

$$\max_{\mathsf{X}^l \in \mathcal{S}'_n} \beta_\alpha(P_{\mathbb{H}^l\mathbb{Y}^l \,|\, \mathbb{X}^l = \mathsf{X}^l}, Q_{\mathbb{H}^l\mathbb{Y}^l}) \gg \beta_\alpha(P_{\mathbb{X}^l\mathbb{H}^l\mathbb{Y}^l}, P_{\mathbb{X}^l}Q_{\mathbb{H}^l\mathbb{Y}^l}) \tag{52}$$

which means that the resulting bound is much looser than (49).

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[2] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[3] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[4] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[5] R. E. Blahut, "Computation of channel capacity and rate-distortion function," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[6] R. G. Gallager, "Source coding with side information and universal coding," 1979. [Online]. Available: http://web.mit.edu/gallager/www/papers/paper5.pdf

[7] S. Shamai (Shitz) and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[8] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, Jan. 2014.

[9] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Trans. Royal Soc. A*, vol. 231, pp. 289–337, 1933.

[10] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[11] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[12] S. Verdú, "The exponential distribution in information theory," *Probl. Inf. Transm.*, vol. 32, no. 1, pp. 86–95, 1996.

[13] T. J. Riedl, T. P. Coleman, and A. C. Singer, "Finite block-length achievable rates for queuing timing channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Paraty, Oct. 2011, pp. 200–204.

[14] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[15] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2959–2963.

[16] C. E. Shannon, "Certain results in the coding theory for noisy channels," *Inf. Contr.*, vol. 1, pp. 6–25, 1957.

[17] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jul. 2009.

[18] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "A beta-beta achievability bound with applications (extended version)," 2016. [Online]. Available: http://arxiv.org/abs/1601.05880

[19] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, 1968.

[20] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2665, May 2014.

[21] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[22] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4880–4902, Aug. 2011.

[23] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. 48th Allerton Conf. Commun., Contr., Comp.*, Monticello, IL, USA, Sep. 2010.